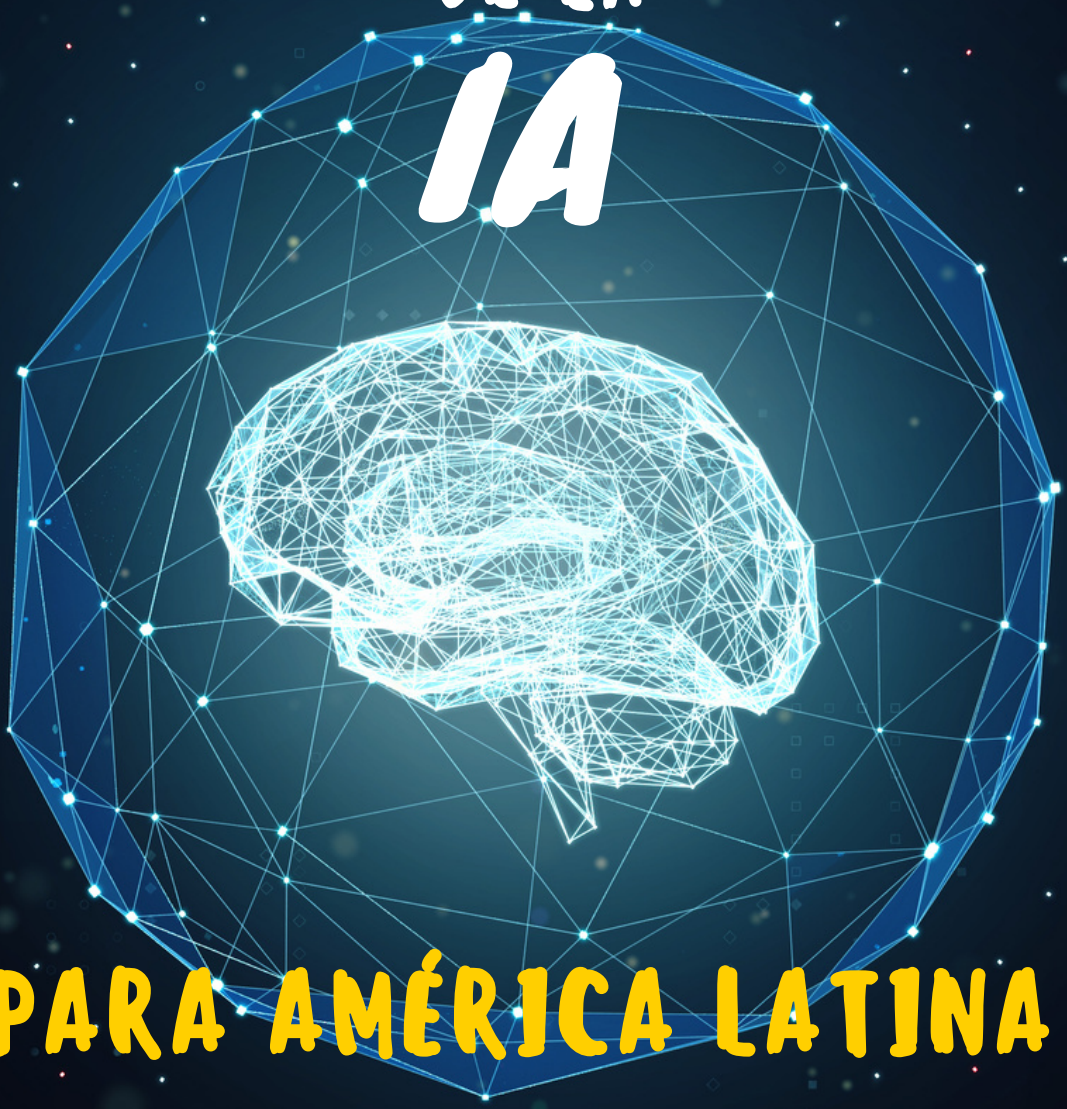
ÉTICA DE LA IA



PARA AMÉRICA LATINA

Un booklet para emprendimientos y empresas desarrollado por:

C MINDS

eon resilience lab by C Minds

Con el apoyo y contribuciones de

Meta

ÉTICA DE LA IA PARA EMPRESAS LATINOAMERICANAS

UN BOOKLET PARA EMPRENDIMIENTOS Y EMPRESAS DE INTELIGENCIA ARTIFICIAL EN AMÉRICA LATINA DESARROLLADO POR C MINDS CON EL APOYO Y CONTRIBUCIONES DE META

Agosto 2022

Autoras: Claudia May Del Pozo, Directora del Eon Resilience Lab de C Minds y Carla Vázquez Wallach, Consultora Legal en C Minds

Contribuidores principales: Norberto de Andrade, Director de Política y Gobernanza en Meta, Paula Vargas, Directora de Privacidad para América Latina en Meta, Constanza Gómez Mont, Fundadora y Presidenta de C Minds, Lucía Trochez, Directora en C Minds, Cristian Guerrero, Consultor Técnico en C Minds y Ana Victoria Martín del Campo Alcocer, ex-Coordinadora de Proyectos en C Minds, Daniel Castaño, Profesor de Derecho en la Universidad Externado de Colombia

Contribuidores: Alejandro Cobando de Talov, Alexei Stanislawski de Maat.ai, Andres Felipe Montoya Nieta de Nediar, Antonio Henrique Dianin de Project Company, Cuco Vega de Bexi, Eduardo Farina de Blue Messaging, Esteban Gorupizc de Atexto, Erick Carranza y Miguel Ángeles del proyecto IRBin, Erick Estrada de Tooring, Genaro Aldana Chavez de ReMap 4.0, Gimena Olguin de Quick Hit Solutions, Ivan Caballero de Citibeats, José Tomás Arenas de TeleDx, Leticia Ramírez de Drone Domain, Oscar Landivar de Wizdem, Pavel Pichardo de Madison, Pedro Vallejo Castillo de Datlas, Rafael Figueroa de Portal Telemedicina, Rodolfo Rubén Alvarez González de Dyoo, Sebastián García de IDATHA, Sebastían Flores de U-Planner, Valeria Resendez de Xira.

CONTENIDO

INTRODUCCIÓN

De acuerdo con la Organización para la Cooperación y el Desarrollo Económicos (OCDE), la inteligencia artificial (IA) se describe como: *"un sistema computacional que puede, para un determinado conjunto de objetivos definidos por humanos, hacer predicciones y recomendaciones o tomar decisiones que influyen en entornos reales o virtuales. Los sistemas de IA están diseñados para operar con distintos niveles de autonomía".*

Sin duda las soluciones tecnológicas basadas en IA están transformando la prestación de los servicios y han fomentado la creación de nuevos productos en el mercado, diversificando las oportunidades de negocio.

Los beneficios del uso de IA son cada vez más evidentes repercutiendo tanto en la eficiencia de las empresas como en la calidad y confianza de los servicios y productos que reciben los consumidores. De acuerdo con la iniciativa fAIr LAC diseñada por el Banco Inter-Americano de Desarrollo y C Minds, en sus inicios las primeras aplicaciones de IA se centraban en el gobierno electrónico y en mejorar los procesos de gobernanza. Sin embargo, se han empezado a desarrollar iniciativas basadas en el uso de IA para atender problemáticas sociales, como mejorar el acceso a la educación, ofrecer mejores servicios de salud, combatir la pobreza o reducir la desigualdad, entre otros.

En el estudio "La inteligencia artificial al servicio del bien social en América Latina y el Caribe: Panorámica regional e instantáneas de doce países" se estimó que la IA podría aportar hasta un 14% de riqueza adicional a las economías emergentes de América Latina.

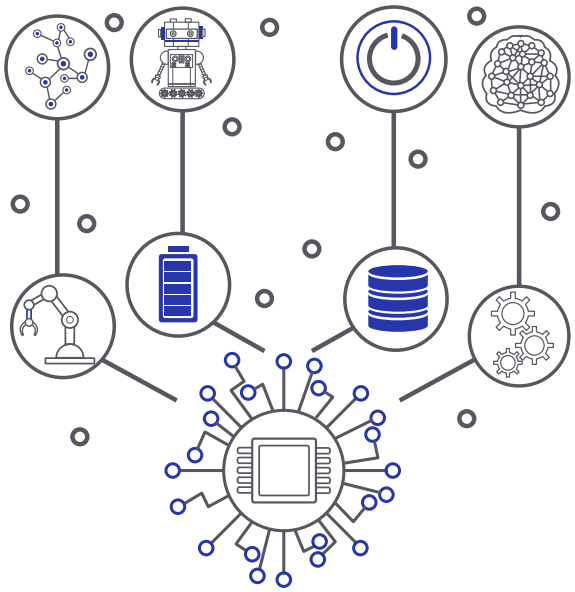
Por su parte la OCDE identificó que el procesamiento masivo de los datos a través del uso de IA, ha permitido atender retos globales mediante la detección de patrones que proporcionan información estratégica a los tomadores de decisiones mejorando su eficiencia.

Sin embargo, el uso de la IA trae consigo múltiples desafíos, particularmente en razón de los posibles riesgos y efectos adversos para la humanidad derivados de la opacidad en el diseño (ver más información en página 10), implementación y uso de sistemas IA. En 2019 el Consejo Ejecutivo de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) reconoció que, si bien la IA tiene potencial para transformar el futuro de la humanidad para mejorar y favorecer el desarrollo sostenible, también existe una conciencia generalizada de los riesgos y desafíos que conlleva, especialmente por lo que respecta a la agravación de las desigualdades y

brechas existentes, así como las implicaciones para los derechos humanos.

En respuesta a esta inquietud, distintas organizaciones y gobiernos se han preocupado por las consideraciones éticas de la IA, desarrollando recomendaciones, directrices, principios, lineamientos, estudios e informes.

A esta preocupación también se ha unido el sector académico y privado, a través del debate y diseño de guías y/o recomendaciones sobre la implementación de conceptos éticos en el uso de IA a través de prácticas concretas que materializan principios y valores en el ciclo de vida del sistema de IA, por ejemplo, la creación de códigos de conducta o políticas internas sobre la implementación de principios éticos en procesos de desarrollo de productos y servicios.



Esta guía fomenta una cultura ética en las personas emprendedoras y las empresas que diseñan, desarrollan e implementan sistemas de IA en América Latina para: (i) lograr un entendimiento y concientización que les permita identificar posibles riesgos y analizar el impacto de sus soluciones, (ii) tomar medidas adecuadas y proporcionales en función de la magnitud de los riesgos identificados, y (iii) cuando proceda prevenir o minimizar riesgos. Bajo el alcance de esta guía, el uso ético de la IA tiene un doble propósito, por un lado, significa poner esta tecnología al servicio de la humanidad, en beneficio de la sociedad y mejorar la vida de los seres humanos y por el otro, reducir riesgos e impactos negativos no deseados que el uso de la IA puede generar, ya sea debido a un diseño, desarrollo o implementación deficiente, o a una aplicación inapropiada.

Es posible que al momento de analizar los impactos de los sistemas de IA se presenten tensiones entre los diferentes principios y no siempre se alcance un equilibrio estricto. Por ello, las empresas deben adoptar un proceso de reflexión continuo sobre las buenas prácticas propuestas en la guía.

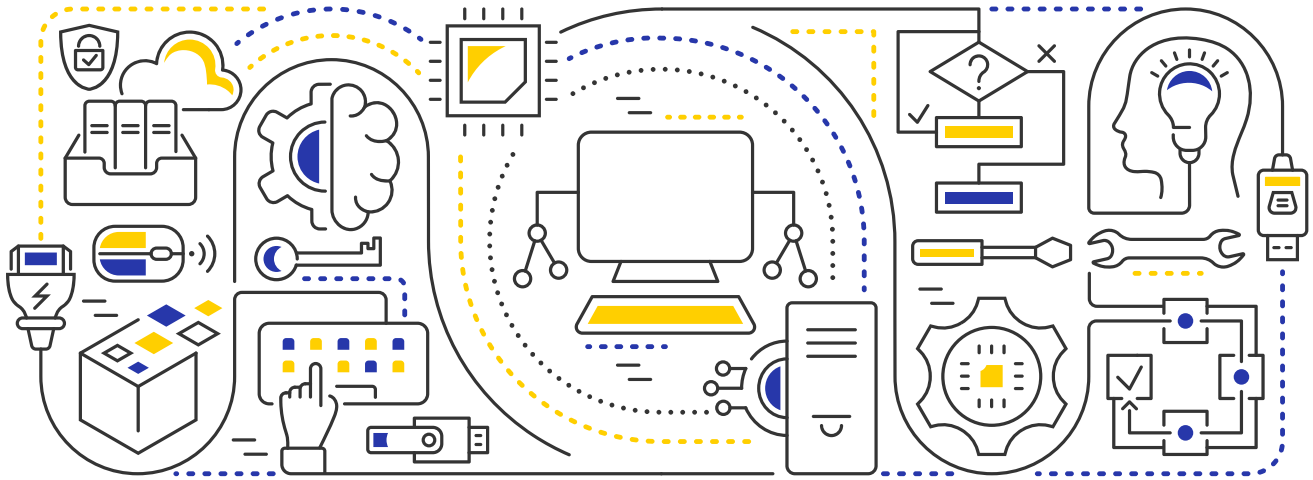
Las empresas deben anticiparse, y de ser necesario, implementar medidas específicas para atender posibles efectos adversos de los sistemas de IA, por ejemplo, diseñar evaluaciones de impacto

periódicas, establecer grados de supervisión y control en la toma de decisiones o robustecer los mecanismos de seguridad, todo ello dependiendo del riesgo de cada sistema, su probabilidad de ocurrencia y posibles consecuencias.

A través de un enfoque práctico, las empresas y los emprendedores podrán adoptar una cultura ética para utilizar sistemas de IA de manera responsable con su entorno, particularmente si dichos sistemas, de alguna manera, impactarán a grupos vulnerables tales como sectores que tradicionalmente se han visto desfavorecidos o discriminados (ya sea por sus características o bien por asimetrías sociales), personas con alguna discapacidad o niños.

La metodología utilizada para el diseño de las buenas prácticas presentadas en esta guía se estructuró con base en:

- El análisis de más de 20 de documentos relevantes (Ver Sección de documentos clave) en materia de ética de IA a nivel internacional;
- La ejecución de entrevistas de diagnóstico con 23 empresas de América Latina (ver anexo 1) que usan sistemas basados en IA para conocer sus perspectivas, necesidades, experiencias y aprendizajes.
- Los principios de la OCDE adoptados por los países miembro en 2019 así como la aprobación de la Recomendación sobre la Ética de la Inteligencia Artificial (IA),



¿PARA QUIÉN ES ESTA GUÍA?

Esta guía se diseñó para empresas y emprendimientos que participan de manera activa en el diseño, desarrollo e implementación de un sistema de IA. Incluyendo todos aquellos que desempeñan un papel activo en el ciclo de vida del sistema de IA, tanto personas físicas como jurídicas.

Para mayor claridad, a continuación se describen los distintos roles que pudieran presentarse en el ciclo de vida de un sistema de IA de acuerdo con UNCITRAL:

Desarrollador(a):
Persona responsable del diseño teórico de alto nivel del sistema de IA, así como de la programación, la capacitación y la verificación de dicho sistema, y de su interfaz e integración con el hardware externo y las aplicaciones y fuentes de datos externos antes de la implantación;

Proveedor(a) de datos:
Persona que proporciona datos o es responsable de que se proporcionen datos al sistema (es decir, los datos necesarios para respaldar la capacitación, la implantación o el funcionamiento);

Implementador(a):
Persona que implementa el sistema integrándolo en sus operaciones (por ejemplo, en los bienes y servicios que suministra), en particular configurando, administrando, manteniendo y respaldando el suministro de los datos y la infraestructura necesarios para el funcionamiento y la supervisión del sistema de IA y su interacción con los datos suministrados una vez implantado;

Operador(a):
persona que hace funcionar el sistema: i) en muchos casos, el operador es la persona que implementa el sistema; ii) en algunos casos, el operador puede ser el usuario final de los bienes o servicios con IA incorporada (por ejemplo, si el usuario final tiene algún grado de control sobre el funcionamiento de esos bienes o servicios); y

Usuario(a) final:
Cualquier otra persona afectada por el funcionamiento de un sistema de IA, incluso al interactuar con el sistema (por ejemplo cuando proporciona datos al sistema) o por ser el usuario final de bienes o servicios con IA incorporada.

La guía será de utilidad para las empresas y emprendimientos que actúan como diseñadores, desarrolladores e implementadores. Cada uno de estos participantes podrá adoptar valores y principios éticos de manera voluntaria para crear una cultura de conciencia y compromiso dentro de su organización y para la sociedad, con dos objetivos específicos:

1. Fomentar la confianza de los usuarios en el uso de inteligencia artificial, incluyendo aquellos que adquieren sistemas de IA para prestar servicios u ofrecer bienes a terceros, y;
2. Mitigar posibles riesgos y efectos negativos en la sociedad.

¿CÓMO USAR ESTA GUÍA?

Esta guía es un instrumento para concientizar a empresas y emprendimientos sobre la importancia y necesidad de implementar principios éticos en el ciclo de vida de un sistema de IA. El grado de adopción de las buenas prácticas contempladas en la guía lo determina la empresa o el emprendimiento de acuerdo con su compromiso ético con el entorno y con los usuarios finales de los bienes o servicios con IA que ofrece, así como a los principios específicos que le apliquen al sistema de IA bajo el contexto en el cual fueron diseñados, desarrollados y/o implementados.

A través de la guía las empresas obtendrán información clara y sencilla para conocer el significado de los principios éticos, así como las buenas prácticas para crear una conducta ética en el uso de sistemas de IA. También podrán medir su avance progresivo mediante la identificación gradual de acciones concretas dentro de su organización y para con sus usuarios.

La guía recoge 5 principios éticos que se describen individualmente en una sección específica con los siguientes rubros:

¿QUÉ ES?

Apartado donde se explica de manera general el Principio para alcanzar un entendimiento claro y sencillo sobre su aplicación en sistemas de IA.

ADOPCIÓN GRADUAL

Contempla buenas prácticas relacionadas con la implementación de cada Principio, así como las actividades a implementar para que las empresas y emprendimientos midan su avance de acuerdo con sus metas y objetivos.

La empresa o emprendimiento podrá implementar las buenas prácticas contenidas en esta guía de acuerdo con el compromiso que vaya adquiriendo sobre los principios éticos. Se incluyen acciones específicas y consecutivas que van reflejando mayor compromiso de la organización sobre el impacto que genera el sistema de IA en su entorno.

Ejemplo:

Equidad, inclusión y no discriminación

Sobre este principio, la guía recomienda 8 niveles, ubicando la número 8 como la acción más compleja en su implementación. Las empresas, en función a su capacidad y recursos, progresivamente decidirán el avance en la adopción.

1. *Revisar la calidad de los datos (incluyendo consistencia, diversidad, integridad, accesibilidad, precisión y completitud).*

2. *Conceptualizar la revisión de los datos como un proceso continuo transversal que va desde el diseño, la implementación técnica de las bases de datos, los estándares y prácticas utilizadas para almacenarlos/modificarlos, así como protocolos de seguridad para accederlos y compartirlos.*

3. *Promover la participación activa de personas con diferentes contextos, sin importar raza, color, ascendencia, edad, género, idioma, religión, opiniones políticas, condición económica o social, en la conceptualización del diseño del sistema de IA e, cuando apropiado, el análisis del impacto del sistema de IA en sus comunidades*

4. *Realizar un análisis de rendimiento para diferentes subgrupos, revisando los efectos que provoca el resultado del procesamiento de datos, así como las decisiones o recomendaciones del sistema de IA.*

¿QUIERES CONOCER MÁS?

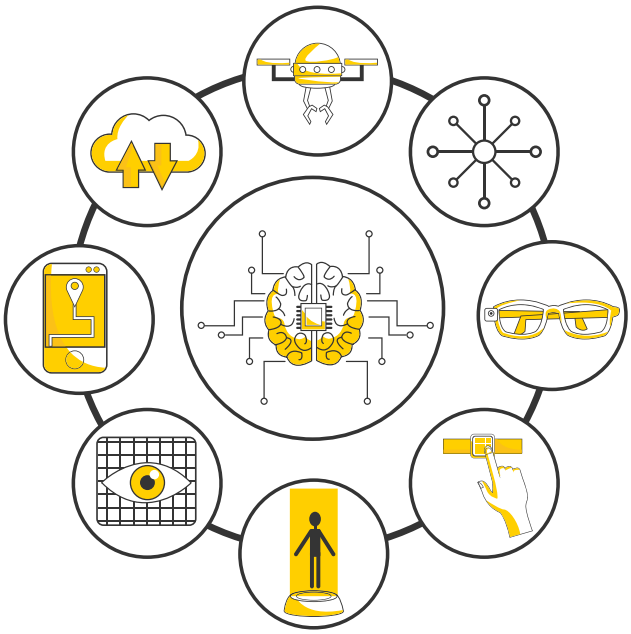
En esta sección se incluyen fuentes adicionales de información para que las empresas y emprendimientos profundicen en el estudio y aplicación de los principios éticos, así como casos de uso y/o ejemplos de la aplicación práctica de dichos principios.

IMPORTANTE

Esta guía no ofrece asesoría legal ni emite recomendaciones sobre el cumplimiento del marco legal aplicable a los sistemas de IA. Tampoco sustituye ninguna normativa legal que aplique al sistema de IA, a la empresa y emprendimiento ni cubre requisitos legales que pudieran ser exigibles al contexto en el que se implementarán sistemas de IA.

La empresa y emprendimiento es la responsable del análisis de factibilidad legal de su modelo de IA, considerando como por ejemplo el marco legal aplicable al caso de uso correspondiente, el respeto a los derechos humanos, propiedad intelectual, protección de datos y privacidad, protección a los derechos del consumidor, requerimientos de calidad o seguridad que pudieran aplicar a productos, requisitos de validez y legalidad de actos emitidos por entes públicos, entre otros.

La guía es un documento dinámico que irá adoptando buenas prácticas para el diseño, desarrollo e implementación responsable de sistemas de IA que en el futuro se identifiquen, así como aquellas que surjan de la reflexión constante para garantizar la protección de la dignidad humana, sus derechos y libertades, por lo que las buenas prácticas contenidas en este documento no deben considerarse como recomendaciones definitivas.



CONTEXTO

¿A QUÉ NOS REFERIMOS CON "INTELIGENCIA ARTIFICIAL"?

Actualmente no existe un consenso sobre el significado de IA. Sin embargo, para fines de esta guía, se adopta la definición de Inteligencia Artificial proporcionada por la OCDE ya que propone una definición pragmática basada en las diferentes etapas del ciclo de vida de un sistema de IA.

La OCDE define a la IA como:

"una máquina que puede, de acuerdo con un conjunto de objetivos definidos por humanos, realizar predicciones, recomendaciones o tomar decisiones que tengan una influencia sobre ambientes reales o virtuales. Los sistemas basados en IA están diseñados para operar con distintos niveles de autonomía."

¿QUÉ ES EL USO ÉTICO DE LA IA?

Las innovaciones en diferentes áreas de la vida moderna han facilitado nuestro día a día mediante el uso de IA , como por ejemplo en el transporte, la comunicación, la medicina, la educación, la ciencia, la vida financiera, el derecho, los servicios de entretenimiento, entre otros. Sin embargo, ha traído consigo debates sobre los retos éticos, que van desde la desaparición de los empleos tradicionales, reproducir y reforzar sesgos existentes, creando brechas de desigualdad más profundas, prejuicios o estereotipos, la responsabilidad por posibles daños físicos o psicológicos a los seres humanos, hasta la deshumanización general de las relaciones humanas y la sociedad en general.

Las empresas y emprendimientos que desarrollan o implementan sistemas de IA deben sensibilizarse sobre los impactos generados en el entorno o en la vida de las personas. Incluso para maximizar los beneficios de la IA eliminando sesgos o discriminaciones que anteriormente no se hubiesen identificado.

CUATRO RAZONES PARA ADOPTAR PRINCIPIOS ÉTICOS EN EL USO DE LA IA

1 **Una oportunidad para generar confianza de la sociedad en la innovación y evolución tecnológica:**

Al adquirir un producto o al recibir un servicio, los usuarios en general tienen la expectativa que sus derechos serán respetados y que sus libertades no se verán comprometidas.

Sin embargo, los sistemas de IA pueden entrañar, incluso de manera no intencionada, repercusiones adversas para los individuos o la sociedad en su conjunto. Cuando las empresas y emprendimientos adoptan una cultura ética de la IA los usuarios confían en el uso de la tecnología no sólo por los beneficios directos del producto o servicio, sino porque perciben el compromiso genuino de la empresa en tomar las precauciones necesarias para mitigar riesgos y efectos negativos hacia ellos.

2 **Una forma de garantizar la sostenibilidad de la compañía:**

Todas las empresas enfrentan distintos retos durante el proceso de maduración de su modelo de negocio.

Sin embargo, para aquellas que incursionaron o incursionarán en el uso de IA, su permanencia dependerá en gran medida de la conciencia social que adopten sobre los posibles impactos que sus productos o servicios provocan en la sociedad pues los usuarios ahora además de considerar la reputación también evalúan lo que una marca dice, hace y representa.

Las empresas que logran vincular con sus consumidores generando conexiones profundas con sus valores obtienen mejores resultados en su rentabilidad.

3 **Una ventaja competitiva y una puerta a nuevos mercados:**

La adopción estructural y voluntaria de valores y principios éticos en el uso de IA se traduce en la creación de ventajas competitivas para las empresas ya que los usuarios desearán interactuar con productos o servicios que se distingan en el mercado por ser responsables, confiables y seguros frente aquellos que no lo demuestren.

En el informe Humanos + Bots: tensión y oportunidad del MIT Technology Review Insights (2018), se menciona que:

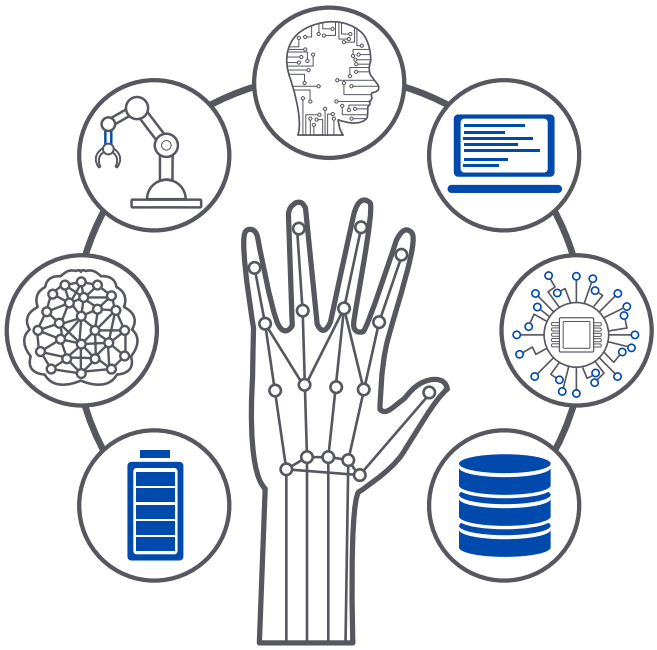
"los cambios de productividad que genera la IA y el aprendizaje automático o

machine learning se traducen rápidamente en poder responder mejor a las diversas demandas de los clientes y satisfacer, de manera consistente, las expectativas que tienen respecto de la marca.

Incluso las empresas que han adoptado una visión ética, por ejemplo en el procesamiento de los datos, no sólo lo ven como un requisito para mantener una posición relevante en el mercado sino como una necesidad de la sociedad en su conjunto.

4 **Mejora el posicionamiento de la empresa para la obtención de financiamiento:**

Las empresas y emprendimientos se enfrentan con frecuencia a desafíos económico-financieros en las fases tempranas de su desarrollo.



Por ello las políticas públicas, particularmente en América Latina, tienen como objetivo impulsar la creación, estabilización y escalamiento de las empresas.

De manera coordinada el sector público y privado han desarrollado redes de apoyo al emprendimiento innovador, tales como grupos de inversionistas ángeles, las incubadoras, aceleradoras, o mecanismos colaborativos de financiamiento.

Particularmente en América Latina, las redes de financiamiento se han enfocado en aquellos emprendimientos que buscan solucionar problemas locales o regionales, por ejemplo, mejorar la eficiencia en la prestación de servicios públicos, reducir la desigualdad social, atender afectaciones al medio ambiente, evitar discriminación a grupos vulnerables, fomentar el acceso a financiamiento, desarrollar capacidades educativas, entre otras.

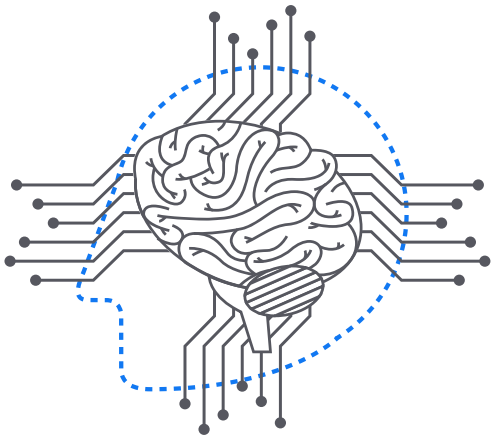
En la evaluación de los proyectos de base tecnológica, particularmente aquellos que usan IA, no sólo se considera el potencial económico sino también la adopción de una cultura empresarial ética para crear valor en la industria o sector con el que interactúa.

RESUMEN DE LOS PRINCIPIOS ÉTICOS DE LA IA

El uso ético de la IA se fundamenta en la voluntad de las empresas y emprendimientos para adoptar una cultura sobre los impactos que dichos sistemas producen en su entorno. A través de los principios éticos se promueve que las empresas y emprendimientos anticipen y lleven a cabo los cambios necesarios para que el uso y aplicación de la IA se realice de manera responsable y ética.

Las buenas prácticas contenidas en este documento plantean una mentalidad de responsabilidad a través de la concientización de principios éticos para los actores que intervienen en el diseño, desarrollo, implementación, monitoreo, supervisión y evaluación de sistemas de IA.

Los principios éticos, a diferencia de las normas obligatorias, no deben interpretarse en sentido estricto, es decir, su adopción es voluntaria y debe analizarse en función de los casos de uso específicos, así como su contexto. Son parte de un proceso iterativo para asegurar que el sistema de IA funciona conforme a los resultados esperados y responde al equilibrio de los intereses involucrados (por ejemplo, el interés de la empresa por desarrollas un sistema con alto potencial de crecimiento que procesa procesa datos sobre el comportamiento de los usuarios en plataformas de entretenimiento (vídeo o música) frente al derecho de privacidad de los usuarios y la expectativa de que el sistema únicamente gestione los datos que resultan indispensables para el uso de dichas plataformas).

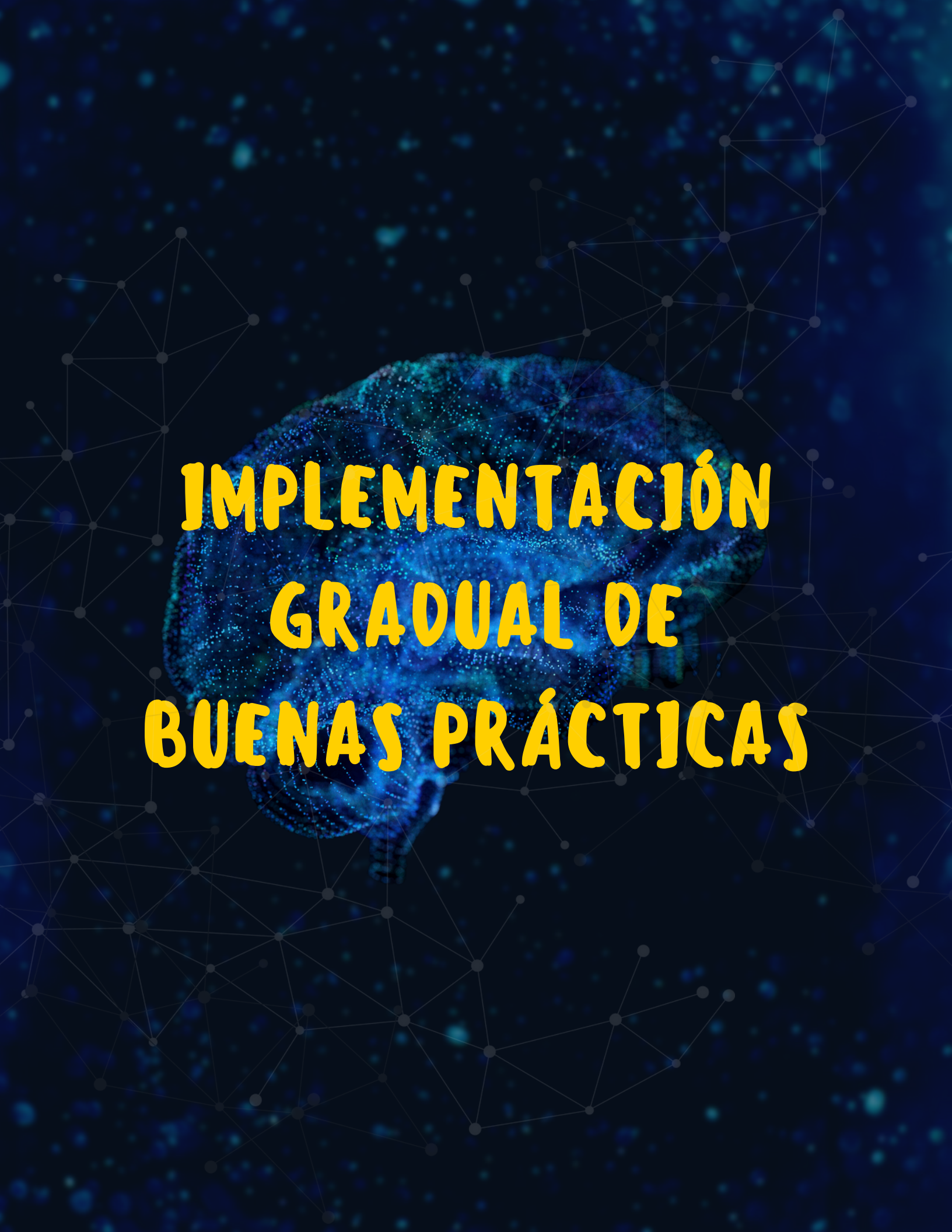


Las empresas y emprendimientos, atendiendo su capacidad y recursos, así como el nivel de compromiso que va adquiriendo sobre los impactos que representa el uso de la IA, voluntariamente deciden progresivamente avanzar en la adopción de los principios éticos de la IA.

La guía recoge cinco principios tomando como referencia las tres construcciones globales más recientes sobre la ética de la IA, la desarrollada por la UNESCO, por la Unión Europea, así como los principios adoptados por la OCDE.

PRINCIPIO	EXPLICACIÓN
 **SEGURIDAD Y ROBUSTEZ**	Igual que cualquier sistema, los sistemas de IA deben asegurar que dicho sistema está diseñado, desarrollado e implementado de manera segura para evitar vulnerabilidades. En atención a los resultados que arroje la evaluación de riesgos de un sistema de IA, será necesario la adopción de estándares de seguridad robustos así como evaluaciones de desempeño periódicas para asegurar que dichos sistemas se comportan de manera segura y funcionan en la forma esperada, incluso cuando pudieran sufrir una ataque o vulnerabilidad.
 **EQUIDAD, INCLUSIÓN Y NO DISCRIMINACIÓN**	La equidad, inclusión y no discriminación, supone garantizar una distribución justa e igualitaria de los beneficios y costes, y asegurar que las personas y grupos no sufran sesgos injustos, discriminación ni estigmatización, buscando que el sistema de IA trate de manera justa a todas las personas usuarias. Los actores de la IA deben desarrollar e implementar sus sistemas de una manera que busque reducir al mínimo daños para el usuario, evitar reforzar o perpetuar los sesgos inadecuados basados en prejuicios o para ciertos grupos de la sociedad, a lo largo del ciclo de vida de los sistemas de IA. Particularmente en aquellos casos que provoquen un daño sustancial, debería disponerse de un recurso efectivo contra la determinación y la discriminación algorítmica injusta.

 **PRIVACIDAD**	Los actores que participan en el ciclo de vida de un sistema de IA deberían fomentar o desarrollar una cultura sobre la gestión de los datos, que abarque no solo el cumplimiento a la protección de los datos personales y privacidad de las personas, sino también un comportamiento responsable sobre el uso de los datos.
 **TRANSPARENCIA Y EXPLICABILIDAD**	La transparencia tiene como objetivo proporcionar información adecuada a los usuarios finales para permitir la comprensión del sistema de IA, fomentar la confianza en el uso de dichos sistemas así como informar los mecanismos de control disponibles sobre las decisiones que le impactan. Las personas tienen derecho a saber cuándo se toma una decisión sobre la base de algoritmos de IA y, en esas circunstancias, contar con mecanismos claros y sencillos para solicitar explicaciones e información sobre ello. La explicabilidad supone hacer inteligibles los resultados de los sistemas de IA y facilitar información sobre ellos.
 **RESPONSABILIDAD Y RENDICIÓN DE CUENTAS**	Aunque un sistema de IA está diseñado para operar de manera autónoma, los humanos tienen un rol relevante en relación con el desarrollo, implementación y uso del sistema. Bajo este principio se identifica que los diferentes actores, en función a su rol, serán responsables del correcto funcionamiento del sistema de IA y deberán rendir cuentas sobre ello, particularmente si se presenta una afectación a los derechos humanos y las libertades fundamentales de los usuarios. Para atender este principio, es necesario aplicar procedimientos o metodologías de evaluación de riesgos durante el ciclo de vida del sistema de IA, integrar recursos efectivos contra daños, mantener una amplia documentación relacionada con el desarrollo, las pruebas y la implementación de sistemas de IA y adoptar medidas para impedir o minimizar los posibles daños.


IMPLEMENTACIÓN
GRADUAL DE
BUENAS PRÁCTICAS

SEGURIDAD Y ROBUSTEZ

SEGURIDAD Y ROBUSTEZ

¿QUÉ ES?

Como cualquier otro sistema, los sistemas de IA también deben adoptar medidas de seguridad para enfrentar posibles vulnerabilidades (ya sea a los datos, el modelo o la infraestructura). De acuerdo con el nivel de madurez del sistema se irán adoptando un conjunto de capacidades clave para fomentar sistemas de IA sostenibles a largo plazo.

Para atender este principio debe desarrollarse un enfoque preventivo de riesgos para asegurar que el sistema sigue funcionando para cumplir el objetivo para el cual se diseñó. Dependiendo del resultado de dicho

análisis, será necesario adoptar medida de seguridad adicionales o más robustas para asegurar que dichos sistemas se comportan de manera segura y funcionan en la forma esperada, incluso cuando pudieran sufrir una ataque o vulnerabilidad.

Este principio debe observarse en todo el ciclo de vida del sistema de IA, para monitorear cualquier cambio significativo que suceda en el entorno con el que sistema interactúe y que pueda requerir - por ejemplo - la inclusión de nuevos datos en el entrenamiento del modelo. La robustez y seguridad de un sistema busca proteger tanto a la empresa como a sus usuarios (directos e indirectos), y su cuidado evitará que la empresa incurra en gastos adicionales para mitigar deudas técnicas o resarcir daños, a su vez previniendo una posible baja de la confianza de los usuarios en el sistema.

ADOPCIÓN GRADUAL

BUENAS PRÁCTICAS

1	Destinar recursos para educar y supervisar la adopción e integración de mejores prácticas de seguridad informática.
2	Implementar una política de control de accesos gradual a los componentes del sistema.
3	Contar con las herramientas adecuadas para lograr un desarrollo ordenado y de calidad, por ejemplo: repositorios de código con control de versiones, mecanismos de revisión de código, documentación suficiente sobre la selección de datos, cómo se desarrolló el modelo, así como las pruebas que se ejecutaron antes de su implementación.
4	Implementar mejores prácticas para la administración de bases de datos, por ejemplo, aislar las bases de datos de cualquier interacción con usuarios o terceros (usando APIs o capas de software que eviten una manipulación directa), la encriptación a nivel sistema de archivos (e.g., Transparent Data Encryption), así como la sanitación de entradas para evitar inyecciones de comandos.
5	Integrar un equipo de aseguramiento de calidad (Quality Assurance, QA) al proceso de desarrollo y despliegue del producto / servicio.
6	Construir software seguro desde el diseño mediante la adopción de mecanismos de seguridad como certificados, encriptación de cualquier transmisión de datos y mejores prácticas de seguridad para reducir vulnerabilidades.
7	Incorporar metodologías de desarrollo de software robusto (como Test-Driven Development (TDD) o Robust Programming).

8	Contar con un modelo de interconexión de sistemas (en inglés Open Systems Interconnection, o OSI) para implementar estrategias y mecanismos de seguridad avanzados.
9	Contar con mecanismos de protección contra ataques.
10	Monitorear el desempeño del sistema de IA a través de métricas y sistemas de alerta automática para lograr una mayor eficiencia de ingeniería (sustentabilidad del modelo).
11	Ejecutar análisis de desempeño sobre largos periodos de tiempo, de forma que se pueda detectar la degradación de la estabilidad del algoritmo o la necesidad de re-entrar el modelo o en su caso establecer mecanismos para incrementar la agilidad del modelo.
12	Definir indicadores de desempeño para evaluar el modelo, por ejemplo, resultados sobre el comportamiento del algoritmo o rendimiento del sistema.
13	Implementar mecanismos que alerten al equipo cuando dichos indicadores tienen comportamiento anómalo.
14	Crear un protocolo de respuesta a emergencias de seguridad y/o fallo en la operación del sistema basado en IA.
15	Integrar equipos de DevOps y SecOps.

¿QUIERES CONOCER MÁS?

El Banco Interamericano de Desarrollo en conjunto con la Organización de los Estados Americanos, a través del observatorio de ciberseguridad, publican el informe sobre Ciberseguridad en América Latina:

- Reporte Ciberseguridad 2020, "Riesgos, avances y el camino a seguir en América Latina y el Caribe", https://publications.iadb.org/publications/spanish/document/Reporte-Ciberseguridad-2020-riesgos-avances-y-el-camino-a-seguir-en-America-Latina-y-el-Caribe.pdf

La Comisión Económica para América Latina y el Caribe (CEPAL) por su parte publicó el análisis sobre la relación entre ciberseguridad y la gobernanza corporativa en el contexto de América Latina:

- CEPAL, " Cybersecurity and the role of the Board of Directors in LAtin America and the Caribbean", Héctor J. Lehuedé, septiembre 2020, https://repositorio.cepal.org/bitstream/handle/11362/45988/S2000552_en.pdf?sequence=4&isAllowed=y

La Organización de los Estados Americanos en conjunto con AWS publicaron una herramienta que permite gestionar los riesgos de ciberseguridad de manera flexible y adaptable a la realidad de cualquier organización, sin importar su tamaño o rubro. Bajo la cual propone tres estrategias para su utilización: (i) Revisión básica de prácticas de ciberseguridad; (ii) Creación o mejora de un programa de ciberseguridad; (iii) Comunicación de los equisitos de ciberseguridad a las partes interesadas.

- Organización de los Estados Americanos con conjunto con AWS publicaron el documento "CIBERSEGURIDAD - MARCO NIST- Un abordaje integral de la Ciberseguridad", 2019, https://www.oas.org/es/sms/cicte/docs/OEA-AWS-Marco-NIST-de-Ciberseguridad-ESP.pdf

También la organización Institute of Electrical and Electronics Engineers (IEEE) ha desarrollado diversos estándares sobre sistemas de IA que han contribuido en la aplicación práctica de principios y marcos de seguridad.

- IEEE, "Artificial Intelligence Systems (AIS) Related Standards", https://standards.ieee.org/initiatives/artificial-intelligence-systems/standards.html.

- IEEE 7000™-2021 - IEEE Standard Model Process for Addressing Ethical Concerns During System Design, https://engagestandards.ieee.org/ieee-7000-2021-for-systems-design-ethical-concerns.html?utm_source=ieeesa&utm_medium=aem&utm_campaign=ais-2021

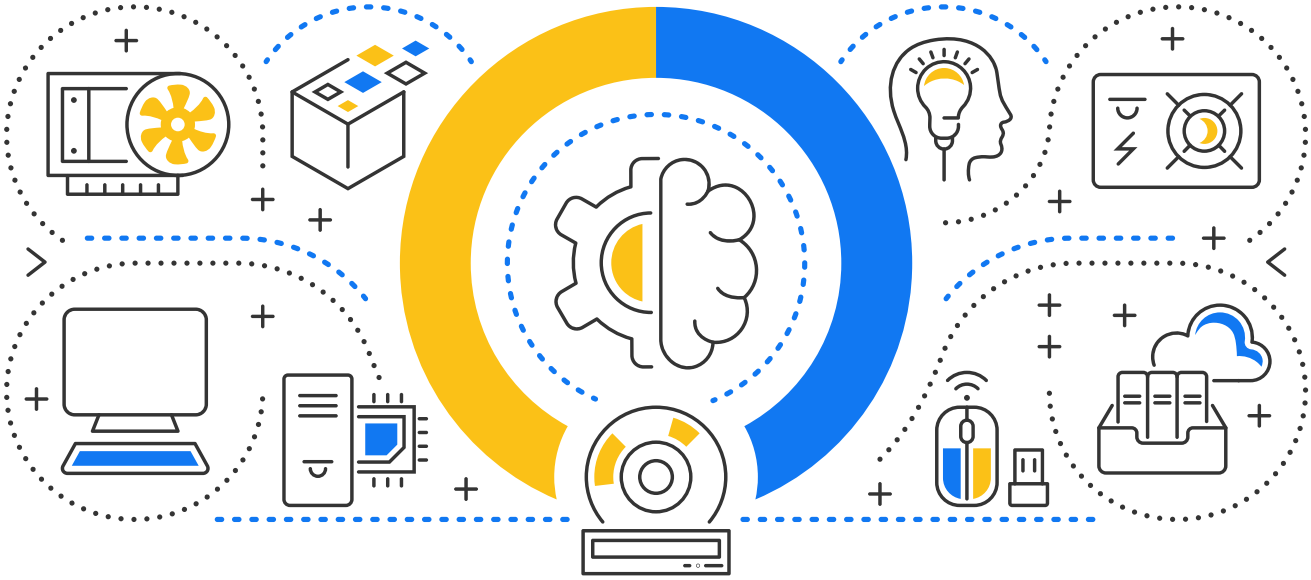
- IEEE P7009™ - Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems, https://standards.ieee.org/project/7009.html.

De igual manera a continuación se mencionan algunos estudios y análisis sobre la IA segura.

- CSET, "CENTER for SECURITY and EMERGING TECHNOLOGY", Tim G.J. Rudner and Helen Toner, "Key Concepts in AI Safety: Robustness and Adversarial Examples", March 2021, https://cset.georgetown.edu/publication/key-concepts-in-ai-safety-robustness-and-adversarial-examples/

- European AI Alliance, "Refining Technical Robustness and Safety Questions", Shahar Avin, November 2019, https://futurium.ec.europa.eu/en/european-ai-alliance/best-practices/refining-technical-robustness-and-safety-questions

EQUIDAD, INCLUSIÓN Y NO DISCRIMINACIÓN



EQUIDAD, INCLUSIÓN Y NO DISCRIMINACIÓN

¿QUÉ ES?

La equidad, un principio estrechamente vinculado con la inclusión y la no discriminación, implica igualdad de trato y oportunidades. Busca que se trate a todos los que interactúen con el sistema de IA de manera justa

Los usuarios de sistemas no deberían sufrir sesgos injustos, estigmatizaciones o estereotipos en función de la raza, color, sexo, idioma, religión, opinión política o de cualquier otra índole, origen nacional o social, posición económica, nacimiento o cualquier otra condición.

El sistema de IA, a través de diseños inclusivos, podría fomentar la accesibilidad de los usuarios para ciertos bienes o servicios, así como su participación. El principio de equidad, inclusión y no discriminación se puede adoptar en tres niveles: (i) en el diseño del sistema; (ii) en las politicas internas de la organización, y (iii) en la implementación. La adopción de este principio facilitará a los programadores de la IA poder identificar si un algoritmo tomará una decisión que pudiera representar un sesgo discriminatorio.

ADOPCIÓN GRADUAL

BUENAS PRÁCTICAS

1	Revisar la calidad de los datos (incluyendo consistencia, diversidad, integridad, accesibilidad, precisión y completitud).
2	Conceptualizar la revisión de los datos como un proceso continuo transversal que va desde el diseño, la implementación técnica de las bases de datos, los estándares y prácticas utilizadas para almacenarlos/modificarlos, así como el diseño de protocolos de seguridad.
3	Promover la participación activa de personas con diferentes contextos, sin importar raza, color, ascendencia, edad, género, idioma, religión, opiniones políticas, condición económica o social, en la conceptualización del diseño del sistema de IA y, cuando apropiado, el análisis del impacto del sistema IA en diferentes comunidades afectadas
4	Realizar un análisis de rendimiento para diferentes subgrupos, revisando los efectos que provoca el resultado del procesamiento de datos.
5	Comparar cómo se comportan métricas para entrenar y evaluar el sistema, a través de los diferentes subgrupos de la población identificados como usuarios y usuarios potenciales. Por ejemplo, el análisis de la tasa de falsos positivos y falsos negativos por subgrupo puede ayudar a la empresa a identificar los subgrupos que experimentan un desempeño desproporcionadamente mejor o peor.
6	Integrar procesos de mitigación de sesgos de manera continua. Las herramientas de evaluación de sesgos pueden ayudar a identificar posibles riesgos de sesgos o incluso mitigar sesgos existentes o futuros.

7 Corregir el sesgo a través de reentrenamiento o enriquecimiento de las bases de datos.

Al detectar patrones discriminatorios que se originen de una base de datos poco representativa, el equipo técnico debería contar con las capacidades para reentrenar el sistema utilizando más y/o mejores datos.

8 Crear un examen que pruebe el sistema para casos frontera o escenarios en los que se prevé que el software pueda comportarse de manera inesperada o fuera del diseño.

¿QUIERES CONOCER MÁS?

Existen distintas herramientas que ayudan a verificar la diversidad en los datos, así como los portales de datos abiertos de diversos países a nivel nacional y local. Estas herramientas también ayudan a comparar los factores que llevan a un algoritmo a tomar una decisión sobre otra.

- OECD, "Tools for Trustworthy AI", A framework to compare implementation tools for trustworthy AI Systems", june 2021, https://www.oecd-ilibrary.org/docserver/008232ec-en.pdf?expires=1644115818&id=id&accname=guest&checksum=1BDB568CCFF115511CD8B68B2BFA7E09
- Banco Interamericano de Desarrollo, fAIr LAC, "Auto-evaluación ética", https://fairlac.iadb.org/es/emprendimiento
- Google, "Visually probe the behavior of trained machine learning models, with minimal coding", https://pair-code.github.io/what-if-tool/.
- META AI, "How we're using Fairness Flow to help build AI that works better for everyone", 2021, https://ai.facebook.com/blog/how-were-using-fairness-flow-to-help-build-ai-that-works-better-for-everyone/

META AI, "How we're using Fairness Flow to help build AI that works better for everyone", 2021, https://ai.facebook.com/blog/how-were-using-fairness-flow-to-help-build-ai-that-works-better-for-everyone/

- META AI, ha desarrollado diversos recursos sobre equidad: https://ai.facebook.com/blog/what-ai-fairness-in-practice-looks-like-at-facebook/

- Microsoft, "AI Fairness Checklist", 2020, https://www.microsoft.com/en-us/research/project/ai-fairness-checklist/.

- ALTAI, "The Assessment List on Trustworthy Artificial Intelligence", https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence

- IBM, "AI Fairness 360", https://aif360.mybluemix.net/.

- IEEE, "IEEE P2863™ - Recommended Practice for Organizational Governance of Artificial Intelligence", 2020, https://standards.ieee.org/project/2863.html.

- IEEE, "IEEE P7003™ - Standard for Algorithmic Bias Considerations", https://standards.ieee.org/project/7003.html.

CASOS DE USO:

- Center for Applied AI at Chicago Booth, "ALGORITHMIC BIAS PLAYBOOK", https://www.ftc.gov/system/files/documents/public_events/1582978/algorithmic-bias-playbook.pdf

PRIVACIDAD

PRIVACIDAD

¿QUÉ ES?

Este principio no pretende sustituir o agotar las obligaciones que las empresas y emprendimientos ya deben atender conforme a los marcos jurídicos nacionales, regionales e internacionales que les correspondan en materia de privacidad y protección de datos.

El principio se refiere a que las empresas integran en el diseño, desarrollo e implementación del sistema de IA, un compromiso ético sobre el uso de los datos y el respeto por la privacidad de las personas usuarias de estos sistemas. El principio contempla la recolección del dato por medio de una autorización y su protección y un uso aceptable y adecuado durante todo el ciclo de vida del dato (recolección, gestión y uso). Se puede fortalecer la privacidad integrándola en el diseño del sistema a través de Tecnologías de Aumento de Privacidad (Privacy Enhancing Technologies, en inglés), que son tecnologías que incorporan los principios fundamentales de la protección de datos al minimizar el uso de los datos personales, maximizar la seguridad de los datos y empoderar a las personas.

Cuando los usuarios interactúan con un sistema de IA depositan su confianza con la expectativa que sus datos serán usados con integridad y honestidad. Esta expectativa se basa en la explicación que se les ha proporcionado sobre cómo serán utilizados sus datos y para qué propósitos. Por ello, las empresas deben balancear sus intereses de negocio con el manejo cuidadoso de los datos de las personas. Entendiendo por este manejo, estándares de debida diligencia en busca del bienestar del titular de los datos y respetando su dignidad.

ADOPCIÓN GRADUAL

BUENAS PRÁCTICAS

1	Reducir, en caso de ser aplicable, la recolección de datos a aquellos que son absolutamente necesarios para los objetivos del sistema de IA y únicamente almacenarlos durante el tiempo necesario. En especial si se trabaja con datos sensibles como datos biométricos o psicométricos.
2	Implementar técnicas de anonimización y pseudo-anonimización a los datos.
3	Asegurar la comprensión profunda de la estrategia de datos en las capacidades de la persona encargada de la supervisión del tratamiento y calidad de los datos.
4	Implementar una política donde los usuarios cuentan con el mayor conocimiento posible sobre cómo sus datos son usados por un sistema de IA (qué datos puede usar un sistema de IA para procesar y cómo debieran ser usados los datos).
5	Cuando sea apropiado, integrar herramientas que permitan al usuario elegir los datos que desea proveer de acuerdo con sus preferencias. Dependiendo del riesgo e impacto del sistema de IA, aclarar que considerando los datos dados, la calidad del servicio o producto podrá cambiar.
6	Estudiar y conocer los alcances de la política de privacidad de terceros (proveedores, colaboradores, aliados).

7	Adoptar la práctica de "privacidad por diseño", la cual implica tener en cuenta el principio de privacidad desde el diseño y a lo largo del proceso de desarrollo, ingeniería y despliegue, integrando el tema de privacidad a la misma tecnología.
8	Realizar evaluaciones de riesgos, considerando el estudio del impacto de privacidad para implementar sistemas efectivos para manejar riesgos y controles internos.

¿QUIERES CONOCER MÁS?

De acuerdo con la Organización de los Estados Americanos (OEA) la mayoría de los países miembros garantizan el respeto y la protección de datos personales como un derecho distinto y complementario a los derechos a la privacidad, la dignidad personal y el honor familiar, la inviolabilidad del hogar y las comunicaciones privadas y conceptos conexos.

La OEA actualizó los principios sobre la privacidad y la protección de datos personales explicando su alcance no solo desde un punto de vista obligatorio sino también de la actitud que se esperaría del responsable del tratamiento de los datos.

- Organización de los Estados Americanos, "Principios actualizados del Comité Jurídico Interamericano sobre la privacidad y la protección de datos personales, con anotaciones", 2021,http://www.oas.org/es/sla/cji/docs/CJI-doc_638-21.pdf.

En 2019, la OEA emitió el documento "Clasificación de los datos", ofreciendo herramientas a las organizaciones para pensar en datos, fundados en la sensibilidad y el impacto comercial, lo que ayuda a la organización a evaluar los riesgos asociados a diferentes tipos de datos.

- Organización de los Estados Americanos & AWS, "Clasificación de los datos", 2019, https://www.oas.org/es/sms/cicte/docs/ESP-Clasificacion-de-Datos.pdf.

META AI ha puesto a disposición OPACUS, una herramienta de código abierto para el entrenamiento de modelos de aprendizaje automatizado con privacidad diferencial para ayudar a

promover el estado de arte y mejorar la privacidad de la IA:

- META AI, "Introducing Opacus: A high-speed library for training PyTorch models with differential privacy", 20201, https://ai.facebook.com/blog/introducing-opacus-a-high-speed-library-for-training-pytorch-models-with-differential-privacy/

La IEEE ha desarrollado estándares relacionados con el ciclo de vida de los datos en sistemas de IA:

- IEEE, "IEEE P2807.1™ - Standard for Technical Requirements and Evaluation of Knowledge Graphs", https://standards.ieee.org/project/2807_1.html.

- IEEE, "IEEE P7002™ - Standard for Data Privacy Process", https://standards.ieee.org/project/7002.html.

- IEEE, "IEEE P7005™ - Standard for Transparent Employer Data Governance", https://standards.ieee.org/project/7005.html.

- IEEE, "IEEE P7006™ - Standard for Personal Data Artificial Intelligence (AI) Agent", https://standards.ieee.org/project/7006.html

La Red Iberoamericana de Protección de Datos emitió recomendaciones bajo un enfoque preventivo para orientar a los desarrolladores sobre las exigencias en el tratamiento de datos personales desde el diseño del producto:

- Red Iberoamericana de Protección de Datos, "Recomendaciones Generales para el Tratamiento de Datos en la Inteligencia Artificial", 2019, https://www.redipd.org/sites/default/files/2020-02/guia-recomendaciones-generales-tratamiento-datos-ia.pdf.

- Red Iberoamericana de Protección de Datos, "Orientaciones específicas para el cumplimiento de los principios y derechos que rigen la protección de los datos personales en los proyectos de inteligencia artificial", 2019, https://www.redipd.org/sites/default/files/2020-02/guia-orientaciones-espec%C3%ADficas-proteccion-datos-ia.pdf.

FAIRsFAIR - aporta soluciones prácticas sobre el uso de los datos a través de la investigación del ciclo de vida de los datos:

- FAIRSFAIR, "Fostering Fair Data Practices in Europe", https://www.fairsfair.eu/tools-software

TRANSPARENCIA Y EXPLICABILIDAD

TRANSPARENCIA Y EXPLICABILIDAD

¿QUÉ ES?

La transparencia y explicabilidad (T&E) no pretende dar acceso completo a los códigos ni modelos de las empresas, pues esto podría afectar sus ventajas competitivas y además no aportaría información clara sobre el funcionamiento de los sistemas. En cambio, la T&E sirve, entre otras cosas, para comprender mejor el alcance de la toma de decisiones automatizadas y las razones de una decisión concreta, así como mejorar el comportamiento futuro de los sistemas.

- La transparencia se refiere a la práctica de hacer visible a los grupos de interés los procesos de sistemas IA, sin que ello implique revelar secretos industrales.

- La explicabilidad es la medida en que la mecánica interna de un sistema de decisión automatizado se puede explicar en términos humanos, conocer las razones que expliquen la toma de una decisión específica por el sistema de IA.

Este principio deberá adoptarse considerando el contexto para el cual el sistema de IA fue diseñado, desarrollado e implementado, pues debe buscarse un equilibrio entre T&E con principios de privacidad y seguridad de los datos.

ADOPCIÓN GRADUAL

BUENAS PRÁCTICAS

1	Informar cuando los usuarios interactúan con IA, generando conciencia al usuario desde los primeros contactos con el sistema de IA, ya sea que las decisiones se basan en algoritmos de IA o que se toma en cuenta determinado resultado de un sistema de IA para a partir de estos para tomar una decisión.
2	Contar con un registro de cómo se diseñó, construyó y mantiene el sistema (tenerlo siempre al día).
3	Atendiendo al contexto, diseñar una comunicación simple evitando lenguaje confuso, de manera clara y adecuada al nivel de alfabetismo digital de los usuarios sobre cómo se implementa cada etapa del sistema de IA para construir conciencia en el uso del sistema de IA . Comunicar el contenido podría responder a preguntas relevantes como por ejemplo:¿Cuál es el nivel de confianza del uso de IA en la decisión que realiza el sistema? ¿Qué datos se usaron para entrenar el modelo?, ¿Qué mecanismo de actualización de datos existe?, ¿De qué manera se garantiza la calidad de los datos usados?, ¿Cuáles son los elementos organizacionales (procesos, modelos de negocio, gobernanza, etc.) ligados a la operación del sistema de IA?, ¿Cuáles son los canales de comunicación entre el usuario y la empresa?, entre otros.
4	Ofrecer información sobre cómo el sistema de IA toma decisiones, dando a conocer los factores que contribuyen a la decisión así como los mecanismos para que los usuarios ofrezcan retroalimentación.

5	Ofrecer mecanismos intuitivos para que los usuarios comprendan cómo funciona el sistema de IA, así como procesos para solicitar explicaciones e información adicional a la proporcionada sobre el funcionamiento del sistema de IA.
6	Cuando la decisión afecte derechos y libertades de los usuarios, facilitar un mecanismo para solicitar revisiones o modificaciones.
7	Dar a conocer la existencia o no de garantías para los usuarios.

¿QUIERES CONOCER MÁS?

Algunas tecnologías de IA ya cuentan con herramientas o bibliotecas provistas por sus creadores o terceros para lograr un mejor entendimiento de cómo funcionan los algoritmos y modelos. En caso de no existir, la explicación de los modelos se convierte en una tarea más avanzada por ser desarrollada por la empresa.

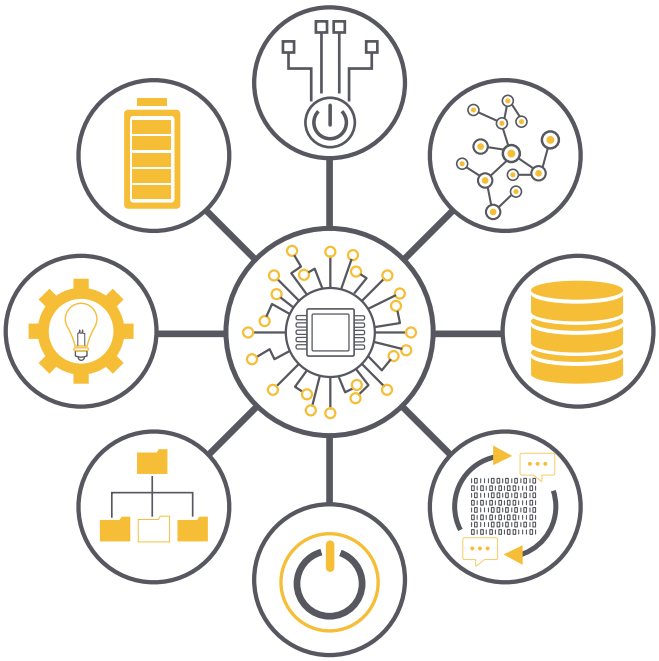
Diversas organizaciones han desarrollado herramientas para orientar a los responsables de la transparencia y explicabilidad de sistemas de IA, aquí algunos ejemplos:

- IBM. "IBM's Principles for Trust and Transparency", https://www.ibm.com/blogs/policy/wp-content/uploads/2018/06/IBM_Principles_SHORT.V4.3.pdf.
- IBM, "AI FactSheets 360", https://aifs360.mybluemix.net/?_ga=2.114415694.891271442.1623897216-1860865525.1623897216&_gac=1.251733499.1623897216.CjwKCAjwwqaGBhBKEiwAMk-FtLnz4b6Inj74bS3CfWK2tIL-RxPxPn_DiqjbXrmgithyEhiQVQ1lTRoCbUoQAvD_BwE.
- IEEE, "IEEE P7001™ - Standards for Transparency of Autonomous Systems", 2020, https://standards.ieee.org/project/2863.html.
- Google, "Responsible AI practices", https://ai.google/responsibilities/responsible-ai-practices/?category=interpretability.

- META AI, "CAPTUM" Model Interpretability for PyTorch, https://captum.ai/

- META AI, "AI Explainability", TTC Labs, https://www.ttclabs.net/theme/ai-explainability

- Microsoft, "InterpretML", https://www.microsoft.com/es-mx/ai/responsible-ai-resources?activetab=pivot1%3aprimaryr4.



- META AI, "CAPTUM" Model Interpretability for PyTorch, https://captum.ai/

- META AI, "AI Explainability", TTC Labs, https://www.ttclabs.net/theme/ai-explainability

- Microsoft, "InterpretML", https://www.microsoft.com/es-mx/ai/responsible-ai-resources?activetab=pivot1%3aprimaryr4.

- Microsoft, "Fairlearn", https://www.microsoft.com/es-mx/ai/responsible-ai-resources?activetab=pivot1%3aprimaryr4.

- Microsoft, "Datasheets for Datasets", https://www.microsoft.com/en-us/research/publication/datasheets-for-datasets/.

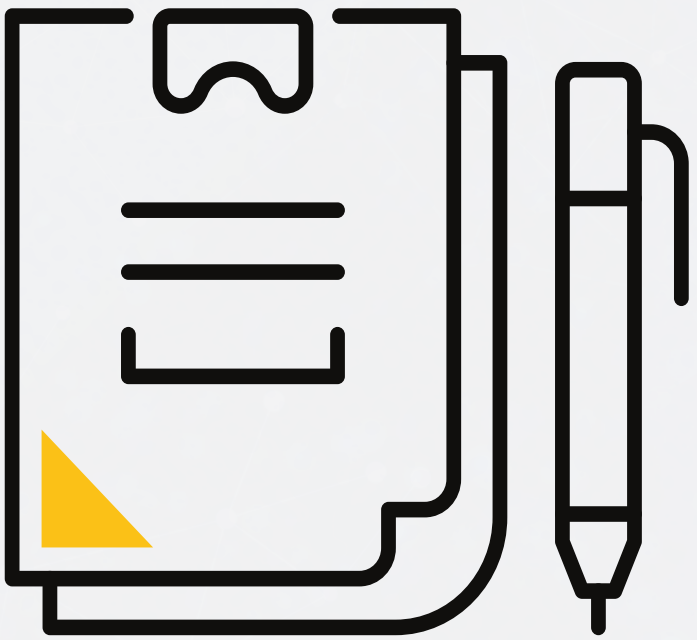
- Oficina del Comisionado de Información del Reino Unido, 2019), "Explaining Decisions with AI. Part 2: Explaining AI in Practice", https://ico.org.uk/media/2616433/explaining-ai-decisions-part-2.pdf.

- TRUST-AI, "Transparent, reliable and unbiased smart tool for AI", https://cordis.europa.eu/project/id/952060

CASOS DE USO:

- Google, "Participatory approaches to transparency in dataset documentation", https://pair-code.github.io/datacardsplaybook/?utm_source=pocket_mylist

RESPONSABILIDAD Y RENDICIÓN DE CUENTAS



RESPONSABILIDAD Y RENDICIÓN DE CUENTAS

¿QUÉ ES?

El sistema de IA debe operar conforme al propósito previamente definido y atender las finalidades sobre las cuales se diseñó, desarrolló, implementó y usa dicho sistema. Las herramientas técnicas ayudan en la identificación y mitigación de los riesgos provocados por un sistema de IA; sin embargo, sólo las personas (actuando de manera individual o grupal) pueden ser responsables por los sistemas de IA.

Por lo tanto, es necesario evaluar el sistema a la luz del contexto específico de su aplicación e identificar posibles riesgos para los usuarios finales, por ejemplo, el sistema de IA que se utiliza como una herramienta que aporta insumos para que un ser humano tome una decisión será distinto de aquel que de forma automatizada toma una decisión sin intervención humana.

Las empresas y emprendimientos establecen la gobernanza del sistema de IA, definiendo políticas internas, así como la responsabilidad que cada uno de los participantes tiene para guiar el diseño, desarrollo, implementación y uso de sistemas de IA dentro de una organización.

ADOPCIÓN GRADUAL

BUENAS PRÁCTICAS

1	Elegir el modelo de IA conforme al objetivo que el sistema de IA busca. Una vez definido el objetivo legítimo que persigue el sistema de IA se debe asegurar que el ciclo de vida del sistema de IA responda adecuadamente a la finalidad determinada (no se deberían ejecutar procesamientos adicionales a lo estrictamente necesario).
2	Tener claridad sobre las responsabilidades de cada persona involucrada en el desarrollo, implementación y uso del sistema de IA.
3	Documentar el desarrollo, pruebas e implementación del modelo del sistema de IA, incluyendo por lo menos las métricas utilizadas, las pruebas ejecutadas para lograr el correcto funcionamiento del sistema, así como el registro trazable y auditable del uso del sistema.
4	Adoptar metodologías de identificación, evaluación y mitigación de riesgos sobre el impacto del sistema de IA durante todo el ciclo de vida del sistema de IA, clasificando el tipo de riesgo, su nivel del impacto y su alcance.
5	Diseñar, en función de la evaluación de riesgos, una estructura de gobierno para tomar decisiones sobre la atención y mitigación de los distintos tipos de riesgos identificados.
6	Incluir el nivel de intervención y supervisión humana en la toma de decisión. Debe considerarse que la supervisión humana es necesaria cuando la decisión afecta directamente a los derechos de una persona o cuando es de alto impacto, esto último en función del resultado que muestre el proceso de evaluación de riesgos).

7	Ofrecer información sobre el modelo adoptado considerando: el propósito y valores que persigue, dar acceso a las políticas que gobiernan el comportamiento, así como las entradas y salidas, cómo funciona el sistema de IA, comportamiento esperado, métodos de entrenamiento, evaluaciones de desempeño realizadas así como limitaciones o riesgos identificadas por ejemplo en qué casos debería o no usarse el sistema de IA, el nivel de efectividad que tiene el modelo (tasa de falsos positivos y negativos) como indicadores de la madurez del sistema y los posibles efectos de su adopción.
8	Contar con mecanismos sencillos y claros para que el usuario pueda solicitar la revisión de una decisión soportada o tomada por un sistema de IA, cuando resulte apropiado y dependiendo de la evaluación de riesgos.
9	Desarrollar un protocolo de respuesta y vigilancia de escenarios imprevistos o indeseados en la operación de sistemas basados en IA, incluyendo recursos efectivos por daños, considerando el análisis y evaluación de riesgos realizado.
10	Crear un consejo interno para la gobernanza de la IA, mediante la cual por ejemplo podrían definirse roles y responsabilidades (de quienes participan en el diseño, desarrollo e implementación del sistema de IA), acciones de verificación, supervisión y cumplimiento, periodicidad de las evaluaciones de riesgos así como flujos de trabajo para escalar la toma de decisiones sobre potenciales daños o mitigación de riesgos.

¿QUIERES CONOCER MÁS?

Existen algunas herramientas que orientan en el proceso de definición de mejores prácticas dentro de la corporación para cumplir con el principio de responsabilidad y rendición de cuentas:

- BID, "IA Responsable Manual técnico - Ciclo de vida de la inteligencia artificial", 2020, https://bit.ly/IAResponsableManualTecnicoCiclodevida

- IBM, "Accountability", https://www.ibm.com/design/ai/ethics/accountability/.

- IEEE, "Ethically Aligned Design", https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf?utm_medium=undefined&utm_source=undefined&utm_campaign=undefined&utm_content=undefined&utm_term=undefined.

- IEEE, "IEEE P2863 - Recommended Practice for Organizational Governance of Artificial Intelligence", https://standards.ieee.org/project/2863.html.

- META AI, "Facebook's five pillars of Responsible AI", 2021, https://ai.facebook.com/blog/facebooks-five-pillars-of-responsible-ai/

- Microsoft, "Responsible AI", https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6.

- OCDE, "Accountability (Principle 1.5)", https://oecd.ai/dashboards/ai-principles/P9.

Asimismo, se listan algunas metodologías para implementar evaluaciones de riesgos atendiendo al contexto en donde se vaya a utilizar el sistema de IA (sector, mercado, usuarios):

- EY, "Supervisory expectations and sound model risk management practices for artificial intelligence and machine learning", https://assets.ey.com/content/dam/ey-sites/ey-com/en_us/topics/banking-and-capital-markets/ey-mrm-ai-ml.pdf?download

- EY, "Understand model risk management for AI and machine learning", 2020, https://www.ey.com/en_us/banking-capital-markets/understand-model-risk-management-for-ai-and-machine-learning.

- McKinsey, "Derisking AI by design: How to build risk management into AI development", 2020, https://mck.co/3yvl73x

- OPEN LOOP, "AI Impact Assessment: A Policy Prototyping Experiment", 2021, https://d32j3j47emgb6f.cloudfront.net/wp-content/uploads/2021/01/AI_Impact_Assessment_A_Policy_Prototyping_Experiment.pdf

- IEEE,"IEEE 7010-2020™ (Standard Now Available) - IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-being", 2020, https://standards.ieee.org/content/ieee-standards/en/standard/7010-2020.html

- PWC, "Model risk management of AI and machine learning systems", 2020, https://www.pwc.co.uk/data-analytics/documents/model-risk-management-of-ai-machine-learning-systems.pdf.

De la mano de los modelos de evaluación de riesgos, es recomendable que la empresa genere una cultura de respeto a los derechos humanos en toda su organización. Para este ejercicio se sugiere consultar el trabajo del Consejo de Europa que ha analizado el impacto de la IA en la esfera de los derechos humanos, democracia y estado de derecho, así como posibles estrategias que pueden implementar las empresas:

- COUNCIL OF EUROPE, AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), " The Impact of Artificial Intelligence on Human Rights, Democracy and the Rule of Law", 2020, https://rm.coe.int/cahai-2020-06-fin-c-muller-the-impact-of-ai-on-human-rights-democracy-/16809ed6da

- COUNCIL OF EUROPE, "Towards Regulation of AI Systems: Global perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe's standards on human rights, democracy and the rule of law.", 2020, https://www.coe.int/en/web/artificial-intelligence/-/-toward-regulation-of-ai-systems

CASOS DE USO:

- Microsoft, Guidelines for Human-AI Interaction, "HAX Toolkit", https://www.microsoft.com/en-us/haxtoolkit/

ANEXO A

Acerca de las empresas participantes:

Atexto (México)

Crea, colecciona y transcribe archivos de voz para ayudar a las máquinas a entender a la gente.

Blue Messaging (México)

Digitaliza la captura de información para procesos en campo sector financiero usando asistentes virtuales.

Citibeats (España)

Ayuda a entender las necesidades de los ciudadanos para que instituciones y gobiernos puedan tomar decisiones más ajustadas.

Bexi (México)

Es una plataforma que diseña, lanza y optimiza campañas de marketing digital.

IRBin de Cirsys y la Pontificia Universidad Católica (Perú)

IRBin clasifica automáticamente la basura para reciclaje a través de reconocimiento de formas y sonidos.

Datlas (México)

Desarrolla soluciones para transformar datos en decisiones.

Drone Domain (México)

Ofrece tecnología para lograr mejores rendimientos, mejor calidad y reducir el impacto ambiental de la agricultura.

DYOO (México)

Desarrolla tecnología de reconocimiento facial para brindar análisis en tiempo real y generar soluciones en seguridad, control de acceso y analítica para análisis de mercado.

Maat.ai (México)

Es una herramienta que permite a las personas crear y compartir su identidad digital.

Nediar (Colombia)

Provee tecnologías de realidad virtual y realidad aumentada, mejorando los procesos de formación por la experiencia, la motivación y el desempeño de aprendizaje.

R1T1 (Brasil)

Es una tecnología de asistencia médica en todas las áreas del hospital

IDATHA (Uruguay)

Permite la detección de bots y noticias falsas.

Madison (República Dominicana)

Es un software de facturación y control de inventario.

Portal Telemedicina (Brasil)

Se enfoca en asistencia médica y diagnósticos.

Quick Hit Solutions (México)

Optimiza la información de las empresas.

REMAP4.0 (México)

Genera rutas inteligentes y fomenta el comercio de proximidad.

Talov (Ecuador)

Traduce de lenguaje de señas para sordomudos a través de video o imágen.

Tooring (México)

Desarrolla sistemas de IA en diferentes disciplinas, incluyendo la geográfica, el sector salud y otros.

Wizdem (México)

Ofrece un sistema de recomendación con proyectos sociales y herramientas de gestión para ayudar a producir proyectos sociales más eficientes.

TeleDx (Chile)

A través de su sistema DART, ofrece diagnósticos médicos a través reconocimiento de imágenes y predicciones de enfermedades

U-Planner (Chile)

Realiza diagnósticos de mejora estudiantil personalizada y guía institucional.

Xira (México)

Automatiza la atención a clientes a través de procesos con Chatbots y RPA.

C MINDS

eon resilience lab by C Minds

Con el apoyo y contribuciones de

Meta